METHODS FOR ALTERING THE CLEAVAGE SPECIFICITY OF A TYPE IIG RESTRICTION ENDONUCLEASE

CROSS REFERENCE

This application is a continuation-in-part of U.S. patent application 10/150,028 filed May 17, 2002, which is a divisional application of U.S. application 09/693,146 filed October 20, 2000, now U.S. Patent No. 6,413,758 issued July 2, 2002, each of which is hereby incorporated by reference in its entirety.

BACKGROUND OF THE INVENTION

Type II restriction endonucleases are a class of enzymes that occur naturally in bacteria and in some viruses. When they are purified away from other bacterial proteins, restriction endonucleases can be used in the laboratory to cleave DNA molecules into small fragments for molecular cloning and gene characterization.

Restriction endonucleases act by recognizing and binding to particular sequences of nucleotides (the 'recognition sequence') along the DNA molecule. Once bound, they cleave the molecule within, to one side of, or to both sides of the recognition sequence. Different restriction endonucleases have affinity for different recognition sequences.

Bacteria that produce restriction endonucleases, protect their own DNA by methylating nucleotides at the cleavage site of the endonuclease. The coordinated production of a restriction endonuclease and a specific methylase is called a restriction-modification (R-M) system.

Methyltransferases are complementary to restriction endonucleases and they provide the means by which bacteria are able to protect their own DNA and distinguish it from foreign, infecting DNA. Modification methylases recognize and bind to the same recognition sequence as the corresponding restriction endonuclease, but instead of cleaving the DNA, they chemically modify one particular nucleotide within the sequence by the addition of a methyl group (C5-methyl cytosine, N4-methyl cytosine, or N6-methyl adenine). Following methylation, the recognition sequence is no longer cleaved by the cognate restriction endonuclease. The DNA of a bacterial cell is always fully modified by the activity of its modification methylase. It is therefore completely insensitive to the presence of the endogenous restriction endonuclease.

By means of recombinant DNA technology, it is now possible to clone genes and overproduce the enzymes in large quantities. Restriction endonucleases have highly specific recognition and cleavage sites. It would be desirable to expand the repertoire of restriction endonucleases by

retaining a particular cleavage position while modifying the recognition site of an enzyme.

SUMMARY OF THE INVENTION

In a preferred embodiment, a method is provided for altering the cleavage specificity of a Type IIG restriction endonuclease where the Type IIG restriction endonuclease is characterized by a cleavage domain adjacent to a methylase domain, the methylase domain located adjacent to a specificity domain. The method includes ligating a first DNA sequence and a second DNA sequence to form a fusion DNA. The first DNA sequence includes a DNA segment encoding a catalytic domain and an N-terminal portion of a methylase domain of a first Type IIG restriction endonuclease. The second DNA sequence includes a DNA segment encoding a specificity domain and a C-terminal portion of a methylase domain of a second Type IIG restriction endonuclease, such that the ligation occurs between sequences encoding the methylase domain. A preparation of host cells are then transformed with the fusion DNA for expressing a Type IIG restriction endonuclease with altered cleavage specificity.

In an additional embodiment, the method described above further includes introducing a mutation into the cleavage domain to enhance the viability of the transformed host cell.

In an additional embodiment, a method described above further includes a sequence corresponding to the N-terminal portion of the methylase which terminates within a methylase conserved motif selected from motifs X, I, II, III, IV, V, VI, VII and VIII.

In an additional embodiment, a method described above further includes a sequence corresponding to the C-terminal portion of the methylase which terminates in a methylase conserved motif selected from motifs X, I, II, III, IV, V, VI, VII and VIII. In preferred embodiments, the N-terminal portion and the C-terminal portion of the methylase are non-overlapping.

In a particular example of the above, the sequence corresponding to the N-terminal portion of the methylase motif terminates between the amino acid sequence encoding motif III and amino acids-NPPY in motif IV.

In one embodiment, ligation occurs by means of a linker sequence attached to the N-terminal portion of the methylase domain and the C-terminal portion of the methylase domain on the first and second DNA segment.

In an embodiment, the fusion DNA encodes an active methylase domain.

In an embodiment, the first and second Type IIG endonucleases are endonucleases with defined cleavage and recognition sites or alternatively, the first Type IIG endonuclease is an endonuclease with defined cleavage and recognition sites and the second Type IIG endonuclease is characterized by a bioinformatic search of a microbial sequence database.

In one embodiment of the invention, a method is provided for forming a non-natural, functional Type IIG restriction endonuclease, wherein the Type IIG restriction endonuclease is characterized by a functional cleavage domain, a functional methylase domain and an altered functional specificity domain compared with a natural form of the functional Type IIG endonuclease. The method includes (a) inserting into a DNA encoding the methylase domain or the specificity domain of the natural form of the functional Type IIG endonuclease, a mutation or a nucleic acid linker sequence for inactivating optionally the cleavage domain; and inactivating (i) the functional methylase domain and the specificity domain or (ii) the functional methylase domain or the functional specificity domain; (b) ligating to the DNA at the mutation or at the linker, a DNA encoding (i) a portion of the methylase and specificity domain or (ii) a portion of the methylase or specificity domain to form a fusion DNA; and (c) transforming a host cell having a marker for detecting expression of a colony expressing a non-natural functional Type IIG restriction endonuclease.

For example, in the above method, the mutation may be positioned within a conserved motif in the methylase domain or the mutation may be a deletion at a 5'-end of the DNA encoding the specificity domain or the mutation is a deletion within the specificity domain. Alternatively, where a linker is utilized, the linker may be a transposon mediated linker insertion sequence. The linker may contain a restriction endonuclease cleavage site which is unique within the DNA encoding the restriction endonuclease.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows a gene organization of *Bpm*I restriction-modification system. Genes *BpmIRM* and *BpmIM1* code for *Bpm*I endonuclease (*Bpm*I endonuclease-methylase fusion protein and *Bpm*I M1, respectively. *Bpm*I-Δ#1, *Bpm*I-Δ#2, and *Bpm*I-Δ#3 are deletion mutants with deletions in the methylation or specificity domains.

Figure 2 shows a DNA sequence of *Bpm*IM1 methylase gene (*BpmIM1*) (SEQ ID NO:1) and its encoded amino acid sequence (SEQ ID NO:2).

Figure 3 shows a DNA sequence of *Bpm*I endonuclease gene (*BpmIRM*) (SEQ ID NO:3) and its encoded amino acid sequence (SEQ ID NO:4).

Figure 4 shows a recombinant *Bpm*I endonuclease activity in column fractions following heparin Sepharose chromatography (Amersham Biosciences, Piscataway, NJ). Lane 1: purified native *Bpm*I endonuclease; lanes 2 to 23: heparin Sepharose column fractions. Fractions 11 to 14 gave rise to complete *Bpm*I digestion of λ DNA. The remaining fractions contain no or partial *Bpm*I activity. Lane 24: 1 kb DNA size marker.

Figure 5 shows a functional domain organization of BpmI R-M fusion protein. The conserved amino acid motifs I and IV are indicated. The linear order of the conserved aa motifs in the γ–type methylase is motifs X-I-II-III-IV-V-VI-VII-VIII. The specificity domain is shared between the endonuclease and methylase activites. However, some amino acid residues may play roles in both specific binding (specificity determinant) and catalysis.

Figure 6 shows a cleavage-deficient and methylation-proficient BpmI variants. Lane 1, protein size marker; lane 2, uncut plasmid DNA (pET28-BpmIRM); lanes 3-20, BpmI variants. Lanes 3, 5-9, 12-14, 17, 20, R^-M^+ variants. Note, the R-M fusion methylase (M1) only conferred partial resistance to the plasmid. Full resistance requires a second methylase, M.BpmI (M2).

Figure 7 shows AcuI deletion variant D80A/D(520-1000) in SDS-PAGE gel in the presence and absence of IPTG. M,

protein size marker (in kDa); lanes 1, 3, 5, 7, 9, 11, IPTG-induced cell extracts; lanes 2, 4, 6, 8, 10, 12, non-induced cell extracts. The predicted molecular weight of the deletion variant is 59 kDa.

Figure 8 shows *in vivo* SOS induction assay (endo-blue assay) for the BpmI/BsgI chimeric enzyme on X-gal plates. Blue colony transformants were re-streaked on X-gal plates to test the stability of blue phenotype. First line streak, "\" corresponds to low activity BstYI mutants (blue colonies, positive control); second line streak, "-" corresponds to white colonies, vector control; third, fourth, and fifth streak lines, blue colonies of BpmI/BsgI chimeric clones. Sixteen clones remained blue while eight clones turn white or partial blue.

DETAILED DESCRIPTION OF THE EMBODIMENTS

The methods described herein address the problem of how to modify the specificity of any restriction endonuclease by genetic manipulation. A chimeric endonuclease is formed in which the specificity domain of a particular Type IIG restriction endonuclease is altered by substitution of part or all of the specificity and/or methylase domains with a complementary portion of the specificity and methylase domains from a second Type IIG restriction endonuclease or a methylase. The complementary portion may be selected from any Type IIG restriction endonuclease or a methylase identified in REBASE

or identified by bioinformatic techniques described in U.S. Patent No. 6,383,770.

The molecular architecture of Type IIG enzymes is composed of three functional domains, the catalytic, methylase, and specificity domains (R^+-M^+-S^+) (Figure 1). These domains are generally aligned from the N-terminal end to the C-terminal end in the order of catalytic domain, methylase domain and specificity domain. Whereas the methylase domain contains highly conserved regions (nine motifs have been identified), the catalytic domain and the specificity domain are generally highly variable between different Type IIG restriction endonucleases. Examples of Type IIG enzymes with established recognition and cleavage sites include *Bpm*I, *Acu*I and *Bsg*I.

Particular examples are provided below to demonstrate how functional chimeric endonucleases can be formed having altered specificity. The methods described are not intended to be limited to these examples but rather are applicable to any Type IIG restriction endonuclease characterized by a single specificity domain for both a methylase and catalytic domain. Accordingly, alteration of the specificity domain or switching specificity domains between different Type IIG restriction endonucleases or between a Type IIG endonuclease and a methylase results in alteration of the specificity domain of both the methylase and the restriction endonuclease domain in the target Type IIG endonuclease. *In vivo* assays to

determine whether a functionally active restriction endonuclease with altered specificity has been successfully produced is described in Example 8 using an *E. coli* strain carrying the *dinD:lacZ* fusion (see U.S. Patent No. 5,498,535 herein incorporated by reference). This assay permits 10,000 colonies to be screened on one plate with the positive colonies appearing blue and the negative colonies, white (Figure 8). This *in vivo* assay avoids time-consuming analysis of individual transformed colonies.

The alteration of specificity of a Type IIG restriction endonucleases described herein results from any of two distinct approaches.

Linker insertion

In one approach to the above problem, a nucleic acid linker is inserted into DNA encoding the methylase or specificity domain of one Type IIG restriction endonuclease. The linker may be sufficient in length to encode 3-12 amino acids. A DNA encoding a complementary portion of a second Type IIG restriction endonuclease or a portion of a complementary region of a second Type IIG restriction endonuclease or all or part of an independent γ-type methylase (not derived from a Type IIG endonuclease) but containing a specificity region which is ligated to the linker. The chimeric DNA encodes a functional restriction endonuclease with altered specificity. Example 3 describes how a DNA linker coding for up to about 10 amino acids may

be inserted between the coding region for a methylase and the coding region for the restriction endonuclease (in this example, *Bpm*I) such that a second methylase region and a specificity region is added to the linker.

It will be clear to one of ordinary skill in the art, that any restriction endonuclease cleavage site that occurs only once in the DNA encoding the Type IIG endonuclease may be used as an insertion site for a linker and consequently for adding all or part of a complementary methylase/specificity domain. Examples of single cut sites in the *Bpm*IRM gene are AfeI, BspHI, BclI, HindIII, and PacI. The linker insertion may inactivate the endonuclease catalytic activity (R^-M^+) or the methylase activity (R^+M^-) or both activities (R^+M^-). However, subsequent to ligation, the *in vivo* assay described in Example 8 can rapidly distinguish active from inactive transformed colonies.

One approach to facilitate linker insertion uses a drug resistant cassette flanked by convenient restriction sites. Following the introduction of the drug resistant cassette (for example, selection of Km^R colonies), the majority of the cassette is removed by restriction digestion and religation, leaving only 3 to 12 codons in-frame insertions in the gene.

Alternative to a random single cutting site or a specific cutting site for insertion of a linker, linkers may be randomly inserted using a transposon-mediated linker insertion system

(GPS™-LS linker scanning system, New England Biolabs, Inc., Beverly, MA). This system generates insertion of, for example, 15 bp "linkers" (5 amino acids) at random positions throughout the R-M-S gene. The linker scanning mutagenesis is carried out by introduction of a transposon carrying a drug resistance. Following transposon insertion into the target gene carried on a plasmid and drug resistant selection, the majority of the transposon is removed by restriction digestion and ligation. Religation results in a 15 bp insertion (5 amino acids). Protein segments that are tolerant to linker insertion can be identified this way and a DNA segment encoding a novel binding specificity can be inserted afterwards in a manner similar to that described above. Again the *in vivo* assay described in Example 8 is a rapid method for screening thousands of colonies for endonuclease activity. Positive colonies include colonies having $R^+M^-S^+$ or $R^-M^*S^+$ DNA (M* indicates under methylation).

Non-linker dependent techniques

In an approach exemplified in Examples 5-8, mutations in one of the conserved motifs of the methylase domain of the target Type IIG restriction endonuclease are created which act as a site for exchange of DNA encoding the specificity domains and a part or all of the complementary portion of the methylase. These mutations can be introduced by PCR mutagenesis described in Example 6 but mutations may also be introduced according to other methods know to one of ordinary skill in the art. For example, US Patent application,

serial number 10/208,557 herein incorporated by reference describes an alternate method for creating mutations at target sites in a DNA. In Example 6, we show the construction of an *Acu*I deletion variant carrying deletion in the methylase domain and specificity domain for use in formation of chimeric proteins by exchange of domains with a second Type IIG restriction endonuclease or γ-type methylase.

Example 7 describes how naturally occurring mutations can be utilized to generate a restriction endonuclease with altered specificity. In this example 6, ThaIVp is used for forming a chimeric enzyme with altered specificity. In Example 8, a chimera between *Bpm*I and *Bsg*I is described.

An alternative approach to the above non-linker techniques is to construct a restriction endonuclease variant library containing nested C-terminal amino acid deletions. The deletion can be carried out by, for example, nuclease digestion or Bal31 nuclease digestion from the C-terminal coding end. After nuclease treatment, the ends are filled-in by Klenow fragment and then ligated to a new DNA specificity (binding) domain. The library DNA carrying new DNA specificity domain is then screened by DNA binding assays or by *in vivo* SOS-induction assay, or other functional assays.

In any of the above approaches either utilizing linkers or relying on mutations , ligation of DNA segments from different

sources can be achieved using any of the established techniques in the art such as ligase mediated ligation, ligation using single stranded ends. (see USP 6,660,475, US 2003-0194736 A1). In addition, ligation of proteins or peptides may be achieved using the intein mediated techniques described in USP 5,496,714 WO 00/18881 and WO 00/47751.

If the resulting chimeric protein that is produced form cells transformed with ligated chimeric DNA is insoluble due to aberrant folding and forms inclusion body, the protein can be refolded using various denaturing agents and refolded by slow dialysis into suitable buffer conditions.

All references cited herein are incorporated by reference.

EXAMPLES

Example 1: Techniques

PCR, site-directed mutagenesis PCR procedure: PCR conditions are of 94°C for 30 sec, 55°C for 30 sec, 72°C for 30 sec to 3 min for 13-25 cycles with 2 to 4 units of Vent® DNA polymerase in the presence of 2-10 mM $MgCl_2$, dNTP, DNA template, and 1x Thermopol buffer. In some cases, the PCR products were purified from a low-melting agarose gel and treated with b-agarase. The DNA was precipitated with ethanol and salt. After vacuum drying, the DNA was

resuspended in TE buffer and used for template for assembling PCR or for restriction digestion.

Plasmid DNA preparation procedure: Qiagen spin columns were used to prepare plasmid DNA. Cells lysis, protein and cellular DNA denatuation were performed with the addition of P1, P2, and N3 buffers. Clarified supernatant containing plasmid DNA was loaded onto Qiagen spin columns and washed with PB and PE buffers. Plasmid DNA was eluted with 10 mM Tris-HCl buffer.

Transformation procedure: Chemically competent cells were prepared by treatment of exponential phase *E. coli* cells with ice-cold 50 mM $CaCl_2$ for 30 min. Competent cells were mixed with plasmid DNA and incubated on ice for 30 min. After 3-5 min heat treatment at 37°C, an equal volume of LB broth was added. Cells were re-grown in a 37°C incubator for one hours. Transformants were plated on LB agar plates with appropriate antibiotics for plasmid selection. X-gal was included in the plate for the incubation of "endo-blue" indicator strain.

Electroporation procedure: Electro-competent cells were prepared by washing *E. coli* exponential phase cells in 10% ice-cold glycerol twice (500 ml 10% glycerol for cell pellet from 1 L cell culture). After mixing the DNA with 100 ml of competent cells electroporation was carried out under the condition of 1900 V, 200 W, 25 mF, 0.1 cm cuvette. One ml of

LB was added to cells and incubated for 1 hour to amplify the transformants. Transformants were plated on LB agar plates with appropriate antibiotics (Ap, Cm, or Km) for plasmid selection.

Preparation of cell extracts: Cells were cultured overnight in a 37°C shaker, pelleted by low speed centrifugation (1800 g). Cells were resuspended in a sonication buffer (50 mM Tris-HCl, pH 7.8, 10 mM β-mercaptoethanol, 50 mM NaCl). Cell lysis was completed with sonication at output 4, 50%, discontinuous burst 5 times with a small sonication tip. The lysate was clarified by centrifugation at 14000 g at 4°C for 10 min. The supernatant (cell extract) was used for the nicking enzyme assay.

DNA sequencing: An AmpliTaq™ dideoxy terminator sequencing kit (Applied Biosystems, Foster City, CA)was used in the sequencing reactions. DNA sequences, resolved on an automated sequencer ABI373A, were edited and analyzed using the Seqed program (Applied Biosystems, Foster City, CA) and a sequence analysis software package (Accelrys Inc., San Diego, CA).

Example 2: Cloning of *BpmI* restriction-modification system in E. *coli*

1. Preparation of genomic DNA and restriction digestion of genomic DNA.

Genomic DNA is prepared from *Bacillus pumilus* (New England Biolabs, Inc. collection #711, Beverly, MA) by the standard procedure consisting the following steps:

(a) cell lysis by addition of lysozyme (2 mg/ml final), sucrose (1% final), and 50 mM Tris-HCl, pH 8.0;

(b) cell lysis by addition of 10% SDS (final concentration 0.1%);

(c) cell lysis by addition of 1% Triton X-100 and 62 mM EDTA, 50 mM Tris-HCl, pH 8.0;

(d) phenol-$CHCl_3$ extraction of DNA 3 times (equal volume) and $CHCl_3$ extraction one time;

(e) DNA dialysis in 4 liters of TE buffer, change 3x; and

(f) RNA was removed by RNAse A treatment and the genomic DNA was precipitated in ethanol and resupended in TE buffer;

Five µg genomic DNA was digested partially with 2, 1, 0.5, and 0.25 units of *Apo*I (recognition sequence R/AATTY) at 50°C for 30 min. Genomic DNA fragments in the range of 2-10 kb were purified through a 1% low-melting agarose gel. Genomic and pBR322 DNA were also digested with *Aat*II, *Bam*HI, *Cla*I, *Eag*I, *Eco*RI, *Hin*dIII, *Nde*I, *Nhe*I, *Sal*I, and *Sph*I, respectively. Genomic DNA fragments were ligated to pBR322 with compatible ends.

2. Construction of *Apo*I partial genomic DNA library and challenge of library with *Bpm*I.

The *Apo*I partial DNA fragments were ligated to *Eco*RI digested and calf intestinal phosphatase (CIP) (New England Biolabs, Inc., Beverly, MA) treated pBR322 vector. The ligated DNA was dialyzed by drop dialysis on 4 L of distilled water and transferred into *E. coli* RR1 competent cells by electroporation. Ap^R transformants were pooled and amplified. Plasmid DNA was prepared from the overnight cells and challenged with *Bpm*I. Following *Bpm*I digestion, the challenged DNA was transformed into RR1 cells. Ap^R survivors were screened for resistance to *Bpm*I digestion. A total of 36 plasmid mini-preparations were made. Two resistant clones, #18 and #26, were identified to be resistant to *Bpm*I digestion. *Aat*II, *Bam*HI, *Cla*I, *Eag*I, *Eco*RI, *Hin*dIII, *Nde*I, *Nhe*I, *Sal*I, and *Sph*I digested genomic DNA were also ligated to pBR322 with compatible ends and genomic DNA libraries were constructed.

However, no apparent *Bpm*I resistant clones were discovered from these libraries after screening more than 144 clones.

3. Subcloning and DNA sequencing of the resistant clone.

One resistant clone #26 contains an insert of about 3.1 kb. The forward and reverse primers of pUC19 were used to sequence the insert. Three *Apo*I and one *Hin*dIII fragments were gel-purified and subcloned in pUC19 and sequenced. The rest of the insert was sequenced by primer walking. A methylase gene with high homology to amino-methyltransferase (N6-adenine methylase) was found within the insert which was named *Bpm*IM1 gene. The *BpmIM1* gene is 1,650 bp, encoding a 549-amino acid protein with predicted molecular mass of 63,702 daltons.

4. Cloning of *BpmI* restriction endonuclease gene (*BpmIRM*) by inverse PCR.

There is one partial open reading frame upstream of *BpmIM1* gene that has 31% amino acid sequence identity to another restriction enzyme *Eco*57I with similar recognition sequence (*Eco*57I recognition sequence: 5'CTG**A**AG N16/N14; (Janulaitis et al. *Nucl. Acids Res.* 20:6051-6056 (1992)); *Bpm*I recognition sequence: 5'CTG**G**AG N16/N14 (see Rebase). Genomic DNA was digested with restriction enzymes *Ase*I, *Bcl*I, *Hae*II, *Hpa*II, *Mbo*I, *Mse*I, *Nla*III, *Pac*I, and *Tsp*509I. The digested DNA was ligated at a low DNA concentration at 2

µg/ml and then used for inverse PCR amplification of *BpmIR* gene. The sequences of the inverse PCR primers were the following:

5' gtggaaacggaccgtattatggtt 3' (232-34) (SEQ ID NO:5)
5' caccagtaaataacaggttattcc 3' (232-35) (SEQ ID NO:6)

Inverse PCR conditions were 94°C 1 min, 55°C 1 min, 72°C 2 min for 35 cycles. Inverse PCR products were derived from *Hae*III and *Nla*III templates, gel-purified from low-melting agarose and sequenced using primers 232-34 and 35.

The primers for the second round of inverse PCR were the following:

5' ttcgtagcaagtacggtccatatcagt 3' (233-76) (SEQ ID NO:7)
5' ccgtatgtacttgataggaataacctg 3' (233-77) (SEQ ID NO:8)

Genomic DNA was digested with *Ase*I, *Bcl*I, *Bsr*FI, *Bst*NI, *Eco*RI, *Hin*cII, *Hin*dIII, *Hpa*II, *Nco*I, *Pac*I, *Pvu*I, *Taq*I, *Tfi*I, and *Xba*I. The digested DNA was ligated at a low DNA concentration at 2 µg/ml and then used for inverse PCR amplification of *BpmIR* gene. Inverse PCR conditions were 94°C 1 min, 55°C 1 min, 72°C 2 min for 35 cycles. Inverse PCR products were derived from *Ase*I, *Hin*dIII, *Hpa*II, and *Taq*I templates, gel-purified from low-melting agarose and sequenced using primers 233-76 and 77.

The primers for the third round of inverse PCR were the following:

5′ aggaactaagaaagttcatagctg 3′ (234-61) (SEQ ID NO:9)
5′ atgcggtattatataacccaacag 3′ (234-62) (SEQ ID NO:10)

Genomic DNA was digested with *Afl*III, *Bsp*HI, *Bst*NI, *Eco*RI, *Hae*II, *Hin*P1I, *Hha*II, *Hin*dIII, *Sty*I, and *Xmn*I. The digested DNA was ligated at a low DNA concentration at 2 µg/ml and then used for inverse PCR amplification of *BpmIR* gene. Inverse PCR conditions were 94°C 1 min, 55°C 1 min, 72°C 2 min for 35 cycles. Inverse PCR products were derived from *Hin*P1I and *Xmn*I templates, gel-purified from low-melting agarose and sequenced using primers 234-61 and 62.

The primers for the fourth round of inverse PCR were the following:

5′ tgacgtcctcttcacctaattcgg 3′ (235-50) (SEQ ID NO:11)
5′ gagtttgtgaagatagaaccattg 3′ (235-51) (SEQ ID NO:12)

Genomic DNA was digested with *Apo*I, *Bst*BI, *Bst*YI, *Cla*I, *Eco*RI, *Nde*I, *Rsa*I, *Sau*3AI, *Ssp*I, *Taq*I, and *Xmn*I. The digested DNA was ligated at a low DNA concentration at 2 µg/ml and then used for inverse PCR amplification of *BpmIR* gene. Inverse PCR conditions were 94°C 1 min, 55°C 1 min, 72°C 2 min for 35 cycles. Inverse PCR products were derived from *Apo*I, *Cla*I, *Nde*I, *Rsa*I, *Ssp*I, and *Taq*I templates, gel-purified

from low-melting agarose and sequenced using primers 235-50 and 51. The *Cla*I fragment (2.4 kb) further extends upstream of *BpmIRM* gene. The rest of the *Cla*I fragment was sequenced using primer walking.

After four rounds of inverse PCR reactions, an open reading frame of 3,030 bp was found upstream of *Bpm*I M1 methylase gene, which encodes a 1,009-amino acid protein with predicted molecular mass of 116,891 daltons. This is one of the largest restriction enzymes discovered so far. By amino acid sequence comparison of *Bpm*I endonuclease with all known proteins in GenBank protein database, it was discovered that *Bpm*I endonuclease is a fusion of two distinct elements with possible structural domains of restriction-methylation-specificity (R-M-S). This domain organization is analogous to the type I restriction-modification system with three distinct subunits, restriction, methylation, and specificity (R, M, and S). Because *Bpm*I is distinct from other Type IIs restriction enzymes, it is proposed that *Bpm*I belongs to a subgroup of Type II restriction enzymes called Type IIG.

5. Expression of *BpmIM1* gene in *E. coli*.

Two primers are synthesized to amplify *BpmIM1* gene in PCR. The primer sequences are:

forward:

5′ agcggatccggaggtaaataaatgaatcaattaattgaaaatgttaat 3′ (238-177) (SEQ ID NO:13)

reverse:

5′ aaggggcatgcttatacttatttcttcgttctattgtttct 3′ (238-178) (SEQ ID NO:14)

Following digestion with *Bam*HI and *Sph*I, the PCR product was ligated into pACYC184 with the compatible ends. The ligated DNA was transformed into ER2566 competent cells. Cm^R transformants were plated at 37°C overnight. Plasmids with *BpmIM1* gene inserts were tested for resistance to *Bpm*I digestion. Two out of 18 clones showed full resistance to *Bpm*I digestion, indicating efficient *Bpm*I M1 expression in *E. coli* cells and *Bpm*I site modification on the expression plasmid. The host cell ER2566 [pACYC-BpmIM1] was used for expression of *BpmIRM* gene.

*Bpm*I M1 methylase also modifies *Xho*I site. *Xho*I recognition sequence 5′CT**C**GAG3′ is similar to *Bpm*I recognition sequence 5′CT**G**GAG3′ with only one base difference. It is concluded that *Bpm*I M1 methylase may recognize the sequence 5′CTNN**A**G3′ and modify the adenine base to generate N6-adenine in the symmetric recognition sequence.

6. Expression of *BpmIRM* gene in *E. coli* using a T7 expression vector.

Two primers were synthesized to amplify the *BpmIRM* gene. The primer sequences were:

5′ caaggatccggaggtaaataaatgcatataagtgagttagtagataaatac 3′ (247-217) (SEQ ID NO:15)

5′ ttaggatcctcatttttcttctcctaacgccgctgt 3′ (238-182) (SEQ ID NO:16)

The 3,030-bp *BpmIRM* gene was amplified in PCR using Taq DNA polymerase, digested with *Bam*HI and ligated into *Bam*HI-digested T7 expression vectors pAII17 and pET21a. After transformation of the ligated DNA into ER2566 [pACYC-BpmIM1], Ap^R Cm^R transformants were screened for the endonuclease gene insert. Seven out of 72 clones contained the insert with correct orientation. However, no *Bpm*I activity was detected in cell extracts of IPTG-induced cells. This was probably due to mutations introduced during the PCR process.

To reduce the mutation frequency, Deep Vent® DNA polymerase was used in PCR reactions to amplify the 3,030-bp *BpmIRM* gene. The forward primer incorporated an *Xba*I site and its sequence is the following:

5′ caccaatctagaggaggtaaataaatgcatataagtgagttagtagata aatac 3′ (238-181) (SEQ ID NO:17)

PCR was performed using primers 238-181, 238-182, and Deep Vent® DNA polymerase. The PCR conditions were 94°C 5 min for one cycle; 94°C 1 min, 55°C 1.5 min, 72°C 8 min for 20 cycles. The PCR product was purified through a Qiagen spin column and digested with *Bam*HI and *Xba*I and ligated to T7 expression vectors pAII17 and pET21at with compatible ends. Eighteen out of 36 clones contain the correct size insert. Ten ml cell culture for all 18 clones containing inserts were induced with IPTG for 3h and cell extracts were prepared by sonication and assayed for *Bpm*I activity. Clone #4 displayed partial *Bpm*I activity. Because this gene was derived by PCR cloning, the entire *BpmIRM* fusion gene was sequenced on both strands and it was confirmed to be wild-type sequence.

7. Partial purification of recombinant *Bpm*I activity.

Five hundred ml of cell culture was made for the expression clone #4 ER2566 [pACYC-BpmIM1, pET21at-BpmIRM]. The late log cells were induced with IPTG and cell extract (40 ml) containing *Bpm*I was purified through a heparin Sepharose column. Proteins were eluted with a NaCl gradient of 50 mM to 1 M. Fractions 6 to 27 contained the most protein concentration and were assayed for *Bpm*I activity on λ DNA. It was found that fractions 15 to 18 contained the most active *Bpm*I activity (Figure 4). The yield was estimated at 1,800 units of *Bpm*I per gram of wet *E. coli*

cells. The specific activity was estimated at 24,000 units per mg of protein. Proteins from fractions 15 to 18 were analyzed on a SDS-PAGE gel and protein bands were stained with Gelcode blue stain. A protein band corresponding to ~115 kDa was detected on the protein gel, in close agreement with the predicted size of 117 kDa.

The *E. coli* strain ER2566 [pACYC-BpmIM1, pET21at-BpmIRM] has been deposited under the terms and conditions of the Budapest Treaty with the American Type Culture Collection on January 22, 2001 and received Accession No. PTA-2598.

Example 3. Deletion of the methylase portion of *Bpm*I RM fusion protein

Two primers were synthesized to amplify the putative endonuclease domain with deletion of the methylase and specificity domains. The deletion clone thus contains only the R portion and the M and S portions were removed. The forward primer was 238-181 as described above. The reverse primer had the following sequence with a *Xho*I site at the 5' end:

5' tgaaatctcgagttatcctgatccacaacatatatctgctat 3' (244-95) (SEQ ID NO:18)

The deletion junction was in motif I of γ type N6 adenine methylase. The γ type N6 adenine methylases contain

conserved motifs of X, I, II, III, IV, V, VI, VII, VIII. The specificity domain (TRD) is located after motif VIII. The *Bpm*I deletion clone (*Bpm*I-Δ#1) still carried motifs X and part of motif I. The specificity domain after motif VIII was also deleted (the remaining portion is shown in Figure 1).

PCR was performed using primers 238-181 and 244-95 and Taq plus Vent® DNA polymerase (94°C 1 min, 60°C 1 min, and 72°C 1 min for 25 cycles). The PCR product was digested with *Xba*I and *Xho*I and cloned into a T7 expression vector pET21b. Sixteen clones out of 36 screened contained the correct size insert and the cells were induced with IPTG for 3h. Cell extract was prepared by sonication and assayed for *Bpm*I activity on λ DNA. However, no apparent *Bpm*I digestion pattern was detected. Only non-specific nuclease was detected in cell extract, resulting in a smearing of DNA substrate. It was concluded that deletion of the methylase and specificity portion of the *Bpm*IRM fusion protein abolished *Bpm*I restriction activity.

To further confirm the above result, another deletion clone was constructed that deleted methylase motifs IV, V, VI, VII, VIII, and the specificity domain. This *Eco*RI fragment deletion mutant contains 1,521 bp (507 amino acid) deletion at the C-terminus half of the fusion protein (*Bpm*I-Δ#2). IPTG-induced cell extract of this mutant also did not display *Bpm*I endonuclease activity.

To delete the specificity domain referred to as the target-recognizing domain (TRD), a *Hin*dIII fragment of 579 bp (193 amino acid) was deleted from the C-terminus of *Bpm*I RM fusion endonuclease (*Bpm*I-Δ#3). IPTG-induced cell extract of the TRD deletion mutant did not show any *Bpm*I endonuclease activity. However, the mutant protein displayed non-specific nuclease activity. It was concluded that the TRD is also required for *Bpm*I endonuclease activity. Deletion of the TRD may abolish or reduce its DNA binding affinity and specificity. By swapping in other N6 methylase and specificity domains, new enzyme specificities can be created.

Example 4. Generation of new enzyme specificity using *Bpm*I RM fusion protein

Since *Bpm*I endonuclease consists of three domains (R-M-S), it is possible to plug in other methylation-specificity domains to create a new enzyme specificity. The *BpmIRM* fusion gene is cloned in a T7 expression vector as described in Example 1. Plasmid DNA is prepared. The γ type N6 adenine methylases contain conserved motifs of X, I, II, III, IV, V, VI, VII, VIII (Malone T. et al. *J.Mol.Biol* 253:618-632 (1995)). Motifs X through VIII and TRD are deleted and a DNA linker coding for one, 3, 5, 7 and 10 bridging amino acids are inserted with a restriction site, preferably a blunt end restriction site, for example, the *Sma*I site. The length of the DNA linker is sufficient to provide steric space for the introduction of the new M-S domains. DNA coding for other γ type N6 adenine methylases containing motifs of X, I, II, III,

IV, V, VI, VII, VIII and TRD are ligated to the digested blunt site (in frame) of the *Bpm*I deletion clone. The ligated DNA is transformed into a non-T7 expression vector. After the insert is verified, the plasmid containing new methylation-specificity domains is transformed into a T7 expression host and induced with IPTG. Cell extract is assayed on plasmid and phage DNA and analyzed for new restriction activity.

Example 5. Construction of a cleavage deficient variant of BpmI

Modification of the specificity of *Bpm*I, *Acu*I, and *Bsg*I which are all Type IIG restriction enzymes that display both endonuclease and methyltransferase (R-M) activity is achieved as follows:

Optional mutation of the restriction endonuclease was carried out to increase viability of cells transformed with an enzyme having altered specificity.

A two-step PCR mutagenesis was carried out to mutate the Asp74 codon (Asp74 to Ala74) in the catalytic domain.

The PCR primers have the following sequences:

PCR reaction 1:

Forward primer (P1):

5'
CACCAATCTAGAGGAGGTAAATAAATGCATATAAGTGAGTTAGTAGAT
A
AATAC 3' (TCTAGA, XbaI site) (SEQ ID NO:19)

Reverse mutagenic primer (P2):
5' GTTTATACGAAGTGTATAAGCTGGATTTTTCTTTGTCTC 3' (SEQ ID NO:20)

PCR reaction 2:

Forward mutagenic primer (P3):
5' GAGACAAAGAAAAATCCAGCTTATACACTTCGTATAAAC 3' (SEQ ID NO:21)

Reverse primer (P4):
5' TTAGGATCCTCATTTTTCTTCTCCTAACGCCGCTGT 3' (GGATCC, *Bam*HI site) (SEQ ID NO:22)

The N-terminal 300 bp coding sequence was amplified in PCR reaction 1 with the following PCR conditions: 94°C for 5 min, 1 cycle; 94°C for 30 sec, 55°C for 30 sec, 72°C for 30 sec for 20 cycles; 72°C, 7 min for 1 cycle, 4 units of Vent® DNA polymerase. The rest of the coding sequence was amplified in PCR reaction 2 with the following PCR condition: 94°C for 5 min, 1 cycle; 94°C for 30 sec, 55°C for 30 sec, 72°C for 1 min for 20 cycles; 72°C, 7 min for 1 cycle, 4 units of Vent® DNA polymerase. PCR products 1 and 2 were purified from a low-melting agarose gel and used as the template for PCR assembly using primers P1 and P4. The assembly PCR conditions were 94°C for 5 min, 1 cycle; 94°C for 30 sec, 55°C for 30 sec, 72°C for 3 min for 20 cycles; 72°C, 7 min for 1 cycle.

The mutagenized PCR product was purified and digested with *Xba*I and *Bam*HI and cloned into a T7 expression vector pET28a. The phenotype of the resulting *Bpm*I variant should be R^- (cleavage deficient) and M^+ (methylation proficient). After screening 18 plasmids for PCR insert and digestion with *Bpm*I endonuclease, 11 clones were found to be resistant to *Bpm*I digestion (data shown in Figure 2). Endonuclease activity was not detected in any of the mutant extracts prepared from IPTG-induced cells. Because multiple rounds of PCR were performed to generate the R^-M^+ variant D74A, it was necessary to re-sequence the entire gene to confirm that no other mutations were introduced. Six sequencing primers were used to sequence the entire gene. R^-M^+ variant D74A clone #4 carried one additional amino acid change at E1007G. In a separate experiment, it was determined that E1007G substitution was not important to *Bpm*I endonuclease activity. The Asp74 to Ala74 substitution abolished *Bpm*I endonuclease activity.

Using the same PCR mutagenesis strategy, another BpmI R^-M^+ variant E88A was constructed. It was confirmed that E88A is deficient in endonuclease activity and proficient in methylase activity. Both D74A and E88A mutants are useful as the recipient for exchange of a new specificity domain to generate novel enzymes. After a new specificity is confirmed by DNA binding assays, the mutated residues D74A or E88A are changed back to the wild-type residue Asp74 or Glu88 to

restore the endonuclease activity. A non-cognate methylase can be used to protect the host DNA.

Example 6. Construction of a deletion mutant carrying deletions in the methylase domain and specificity domain

The methylase domains of *Bpm*I and *Acu*I belong to the g type N6 adenine methylases. Motif IV is a conserved methylase block and has a GNPPY sequence in both of the *Bpm*I and *Acu*I methylase domains. This site was chosen as a fusion junction for making chimeric enzymes. An *Acu*I deletion mutant was constructed that deleted methylase motif IV and the remaining C-terminal coding sequence. The starting AcuI enzyme was a cleavage-deficient variant D80A (R-M+ mutant). The codon Phe520 was mutated to a stop codon by PCR mutagenesis to generate variant AcuI D80A/Δ(520-1000). The deletion mutant protein was expressed in *E. coli* ER2566 via T7 expression vector pET28a. When the cells were induced with IPTG (3 hours induction at 37°C), a prominent protein band of 59 kDa was detected in SDS-PAGE gel (data shown in Figure 7). The deletion mutant AcuI D80A/Δ (520-1000) is soluble in *E. coli* cell extract and not degraded by *E. coli* proteases. This deletion mutant can be used as the backbone to construct chimeric Type IIG enzymes. DNA coding for similar methylase motifs IV to VIII and an alternate specificity determinant can be ligated to this deletion mutant to construct a functional chimeric enzyme.

The coding sequences are ligated together by T4 DNA ligase (blunt end or sticky end ligation). In an alternative embodiment, the two coding sequences are assembled together by a two-step PCR method as described in Example 7. After the new specificity is determined by DNA binding assays, the catalytic residue Asp80 is restored. Although not required always, a non-cognate methylase is used here for protection of the host cell after transformation with the fusion DNA encoding the chimeric protein..

Example 7. Expression of a natural deletion mutant of ThaIVp, a truncated Type IIG enzyme

ThaIVp is derived from a thermophilic bacterium. The ThaIVp coding sequence was amplified from the genomic DNA of Thermoplasma acidophilum (ATCC #25905) by PCR. The PCR primers have the following sequences:
Forward:
5'GGTGGTTCTAGAGGAGGTAAATAAATGTCTAATGAAAATTATAACATT GATTTC 3' (TCTAGA, XbaI site) (293-282) (SEQ ID NO:23)
Reverse:
5' GGTGGTGAGCTCCTATTGACATAATCGATCATCAAGAAG 3' (GAGCTC SacI site) (293-283) (SEQ ID NO:24)

The PCR components are as following: 293-282, 293-283 (0.8 mM), 4 units of Vent® DNA polymerase, *Thermoplasma acidophilum* (ATCC # 25905) genomic DNA (1 mg), dNTP (4 mM), 1x thermophilic polymerase buffer, H_2O 73 μl, and $MgSO_4$

at 2, 4, and 6 mM. The PCR conditions are 95°C 5 min for 1 cycle; followed by 95°C 30 sec, 57°C 30 sec, 72°C 3 min for 30 cycles. The PCR product was digested by *Xba*I and *Sac*I and ligated to pET-28a with compatible ends and transferred to T7 expression host ER2566 by transformation. There is a natural stop codon at the end of ThaIVp before the conserved methylase motif IV. Therefore the natural deletion mutant of *Tha*IVp is used as the recipient backbone for generating chimeric Type IIG enzymes by addition (ligation) of motif IV and the remaining methylase motifs and the specificity domain. This DNA segment can be ligated to DNA encoding thermophilic methylase domains (IV to X) and a specificity domain to produce a thermostable chimeric enzyme.

Example 8. Construction of a chimeric enzyme between *Bpm*I and *Bsg*I

The DNA recognition sequences for *Bpm*I and *Bsg*I are CTGGAG and GTGCAG, respectively. *Bsg*I endonuclease is a Type IIG enzyme that shares 35.4% amino acid sequence identity to *Bpm*I. A chimeric enzyme was constructed between *Bpm*I and *Bsg*I, in which the N-terminal coding sequence (catalytic domain plus methylase motifs I to III) was derived from *Bpm*I and the C-terminal coding sequence (methylase motifs IV to X and the specificity domain) was derived from *Bsg*I. The chimeric coding sequence was generated by a two-step PCR reaction. PCR primers were designed that can anneal to methylase motif IV on both *Bpm*I and *Bsg*I

templates. The amino acid sequences in the fusion junction are shown below:

*Bpm*I F D A I I G N P P Y

*Bsg*I F D V I L G N P P Y

The forward primer P1 described in Example 5 and a new reverse primer P2' were used to amplify the N-terminal coding sequence from *Bpm*IRM gene.

The new reverse mutagenic primer P2' has the following sequence:

5' ATAGGGTGGATTGCCTAATATTACATCAAAGCCACCATTTGC 3' (P2'). (SEQ ID NO:25)

PCR conditions were 94°C for 5 min, 1 cycle; 94°C for 30 sec, 55°C for 30 sec, 72°C for 2 min for 17-22 cycles; 72°C, 7 min for 1 cycle with 4 units of Vent® DNA polymerase.

The forward mutagenic primer in the fusion junction has the following sequence:

5' TTTGATGTAATATTAGGCAATCCACCCTATATAAGAATTC 3' (P3') (SEQ ID NO:26)

Since the *Bsg*IRM gene was cloned in pUC19, primer P3' and a pUC universal primer NEB #1221 was used to amplify the C-terminal BsgI coding sequence. PCR conditions were 94°C for 5 min, 1 cycle; 94°C for 30 sec, 55°C for 30 sec, 72°C for 2 min for 15-22 cycles; 72°C, 7 min for 1 cycle with 4 units of Vent® DNA polymerase. The PCR products were purified from a low-melting agarose gel and assembled by PCR using primers P1' and pUC universal primer #1221 (New England Biolabs, Inc., Beverly, MA). The PCR conditions were 94°C for 5 min, 1 cycle; 94°C for 30 sec, 55°C for 30 sec, 72°C for 3 min 10 sec for 15

cycles; 72°C, 7 min for 1 cycle with 4 units of Vent® DNA polymerase. The PCR DNA fragment was cloned into pET21at and transformed into T7 expression host ER2566. *E. coli* host with pACYC-*Bpm*IM or pACYC-*Bsg*IM was also used for transformation. The fusion junction was confirmed by DNA sequencing.

E. coli strain ER1992 carries the *dinD*::*lacZ* fusion (the dinD DNA damage inducible promoter is fused to the *lacZ* gene). When bacterial DNA is damaged by double-stranded cuts or single-stranded nicks, UV radiation, or interference with DNA replication, the indicator strain forms blue colony on X-gal plates. When plasmids carrying the chimeric *Bpm*I/*Bsg*I R-M fusion were transformed into the endo-blue indicator strain ER1992 (*dinD*::*lacZ*), they caused formation of blue colonies in the absence of IPTG induction. This indicates transformants suffered DNA damage resulting from constitutive expression of the fusion protein (data shown in Figure 4). The transformants initially formed blue colonies on X-gal plates. When these cells were plated on X-gal plates with IPTG, most cells turn white. After IPTG induction, cells suffered lethal level of DNA damage and died. The cells carrying inactive mutant version of the chimeric R-M fusion protein took over the population and formed the white colonies. When the blue transformants were re-streaked on X-gal plates, about two-third remains blue colonies, and one-third form white or partial blue colonies.